



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 4 (2025) pp: 2359-2366

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Analisis Prediksi Harga Emas Menggunakan Regresi Linear dan K-Nearest Neighbors

Ricky Martin¹, Faisal², Ikhsan Nur Aditya³, Ahmad Jurnaidi Wahidin⁴, Beni Rahmatullah⁵, Ika Kurniawati⁶

^{1,2,3}Informatika, Universitas Bina Sarana Informatika

^{4,5}Teknologi Informasi, Universitas Bina Sarana Informatika

⁶Sistem Informasi, Universitas Nusa Mandiri

¹martinricky556@gmail.com, ²faisal505gt@gmail.com, ³ikhsanadhitya1@gmail.com, ⁴ahmad.ajn@bsi.ac.id,

⁵beni.brh@bsi.ac.id, ⁶ika.iki@nusamandiri.ac.id

Abstrak

Penelitian ini membandingkan kinerja dua algoritma machine learning, yaitu Regresi Linear dan K-Nearest Neighbors (KNN), dalam memprediksi harga emas berdasarkan data harga penutupan harian dari Januari 2014 hingga Januari 2025. Dataset yang digunakan diperoleh dari platform Kaggle dan merepresentasikan dinamika harga emas internasional selama sekitar sepuluh tahun, sehingga dianggap cukup representatif untuk menggambarkan pola pergerakan harga jangka panjang. Proses pra-pemrosesan data meliputi pembersihan data untuk menghapus missing values, normalisasi seluruh fitur numerik menggunakan StandardScaler, serta pembagian data menjadi 80% data latih dan 20% data uji untuk memastikan proses pelatihan dan pengujian model berjalan terukur. Model dikembangkan menggunakan bahasa pemrograman Python dengan pustaka Scikit-learn. Nilai K pada algoritma KNN ditetapkan sebesar 5 berdasarkan hasil eksperimen awal yang mempertimbangkan keseimbangan antara kompleksitas model dan kinerja prediksi. Kinerja model dievaluasi menggunakan tiga metrik utama, yaitu koefisien determinasi (R^2), Root Mean Squared Error (RMSE), dan Mean Absolute Error (MAE). Hasil pengujian menunjukkan bahwa Regresi Linear memberikan performa terbaik dengan nilai R^2 sebesar 0,9996, RMSE 0,0021, dan MAE 0,0008. Sebaliknya, KNN hanya mencapai R^2 sebesar 0,1107, RMSE 0,0834, dan MAE 0,0412. Perbedaan signifikan ini menunjukkan bahwa hubungan antara data historis harga emas dan nilai prediksinya cenderung linier dan dapat dimodelkan secara efektif oleh Regresi Linear. Sementara itu, pendekatan berbasis kedekatan tetangga seperti KNN kurang mampu menangkap pola tersebut secara optimal. Oleh karena itu, Regresi Linear lebih sesuai digunakan untuk prediksi harga emas berbasis data historis pada periode penelitian ini dibandingkan algoritma KNN.

Kata kunci: Prediksi Harga Emas, Regresi Linear, K-Nearest Neighbors, Machine Learning, Deret Waktu

1. Pendahuluan

Harga emas merupakan salah satu indikator penting dalam aktivitas ekonomi global. Emas banyak dimanfaatkan sebagai instrumen investasi, lindung nilai (*hedging*), maupun cadangan devisa oleh bank sentral. Pergerakan harga emas dipengaruhi oleh berbagai faktor, seperti kondisi ekonomi makro, tingkat suku bunga, inflasi, serta dinamika pasar keuangan internasional. Fluktuasi harga yang tinggi menjadikan prediksi harga emas sebagai masalah yang penting bagi investor, pelaku pasar, maupun peneliti di bidang keuangan dan teknologi informasi.

Berbagai penelitian telah menerapkan metode *machine learning* untuk memodelkan dan memprediksi harga emas. Salah satu penelitian menggunakan regresi linear dan regresi polinomial untuk memprediksi harga emas dan menunjukkan bahwa penambahan komponen nonlinier mampu menurunkan kesalahan prediksi pada data historis [1]. Penelitian lain membandingkan beberapa algoritma pembelajaran mesin untuk prediksi harga emas dan melaporkan bahwa model ensemble dapat memberikan performa yang kompetitif pada berbagai kondisi pasar [2]. Terdapat pula ulasan yang membahas peramalan harga emas dengan memasukkan berbagai faktor ekonomi makro ke dalam model, sehingga prediksi tidak hanya bergantung pada harga historis [3]. Penelitian lain menggunakan beberapa teknik *machine learning* dan menekankan pentingnya pemilihan fitur serta pengaturan parameter agar model tidak hanya akurat tetapi juga stabil [4].

Perkembangan teknologi komputasi dan ketersediaan data historis dalam jumlah besar mendorong pemanfaatan metode *machine learning* untuk memodelkan dan memprediksi harga komoditas. Berbagai penelitian telah menerapkan model deret waktu klasik maupun metode *machine learning* untuk memprediksi harga emas, di

antaranya menggunakan regresi linear dan regresi polinomial [1], beberapa model *machine learning* seperti Ridge Regression, Random Forest, dan K-Nearest Neighbors (KNN) [2], serta pendekatan lain yang membandingkan berbagai teknik pemodelan untuk memperoleh model dengan akurasi prediksi terbaik [4]. Penelitian lain secara khusus menerapkan algoritma KNN untuk memprediksi harga emas dunia dan menunjukkan bahwa metode ini dapat menghasilkan tingkat kesalahan yang relatif rendah pada data historis [5]

Meskipun demikian, banyak penelitian sebelumnya lebih berfokus pada pengembangan model yang kompleks atau pada perbandingan beberapa algoritma *machine learning* secara umum [2], [4], sedangkan perbandingan sistematis antara metode yang relatif sederhana, seperti Regresi Linear dan KNN, masih terbatas. Selain itu, belum banyak kajian yang secara eksplisit membahas kinerja dua algoritma tersebut pada data harga emas harian dalam periode yang panjang, dengan prosedur evaluasi yang seragam dan terukur menggunakan metrik kesalahan yang umum digunakan.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk membandingkan kinerja Regresi Linear dan KNN dalam memprediksi harga emas harian menggunakan data historis periode Januari 2014 hingga Januari 2025. Penelitian ini berfokus pada analisis akurasi prediksi kedua model dengan menggunakan koefisien determinasi (R^2), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE). Kebaruan penelitian terletak pada penerapan dan perbandingan dua algoritma yang relatif sederhana namun banyak digunakan, dengan skema pra-pemrosesan data yang terstandardisasi serta periode pengamatan yang cukup panjang.

Secara khusus, tujuan dari penelitian ini adalah: (1) menganalisis kinerja Regresi Linear dalam memprediksi harga emas harian berbasis data historis; (2) menganalisis kinerja algoritma KNN pada data yang sama; dan (3) membandingkan hasil kedua model untuk menentukan algoritma yang paling sesuai digunakan pada kasus prediksi harga emas pada periode penelitian. Hasil penelitian diharapkan dapat memberikan referensi bagi peneliti dan praktisi dalam memilih metode prediksi yang tepat, sekaligus menjadi dasar bagi pengembangan model prediksi yang lebih kompleks pada penelitian selanjutnya.

1.1 Penelitian Terkait

Beberapa penelitian sebelumnya telah membahas prediksi harga emas dengan pendekatan dan algoritma yang beragam. Sebuah penelitian menggunakan regresi linear dan regresi polinomial untuk memodelkan harga emas dan menunjukkan bahwa penambahan komponen nonlinier mampu menurunkan kesalahan prediksi pada data historis [1]. Penelitian lain membandingkan sejumlah algoritma pembelajaran mesin untuk prediksi harga emas dan melaporkan bahwa model ensemble dapat memberikan performa yang kompetitif pada berbagai kondisi pasar [2]. Studi lain mengevaluasi beberapa teknik *machine learning* untuk prediksi harga emas dan menekankan pentingnya pemilihan fitur serta pengaturan parameter agar model tetap stabil pada berbagai periode pengamatan [4]. Terdapat pula ulasan yang mengkaji pendekatan peramalan harga emas dengan mengintegrasikan berbagai faktor ekonomi makro ke dalam model prediksi, sehingga memberikan gambaran yang lebih komprehensif terhadap pergerakan harga emas [3].

Seiring berkembangnya *deep learning*, sejumlah penelitian mulai memanfaatkan arsitektur jaringan saraf untuk memodelkan deret waktu harga emas. Beberapa studi menggunakan model Long Short-Term Memory (LSTM) untuk memprediksi pergerakan harga emas jangka pendek dan menunjukkan bahwa memori jangka panjang pada model tersebut efektif untuk menangkap pola dinamika harga [6], [7], [8]. Penelitian lain mengusulkan penggunaan arsitektur LSTM, jaringan saraf dalam, atau kombinasi model seperti CNN dan CNN-GRU untuk meningkatkan akurasi prediksi, dan menyimpulkan bahwa model nonlinier mampu menangkap pola fluktuasi harga yang sulit direpresentasikan oleh model linier [9], [10]. Selain itu, terdapat studi yang secara khusus membandingkan arsitektur CNN dan GRU untuk prediksi harga emas menggunakan pendekatan *deep learning* dan menunjukkan bahwa pemilihan arsitektur jaringan berpengaruh terhadap akurasi hasil prediksi [11]. Ada juga studi yang membandingkan model deret waktu klasik seperti ARIMA dengan model LSTM untuk prediksi harga emas dan melaporkan bahwa model LSTM memberikan peningkatan akurasi pada beberapa skenario pengujian [12].

Di sisi lain, penggunaan algoritma K-Nearest Neighbors (KNN) dalam prediksi harga emas juga telah dikaji. Salah satu penelitian menerapkan algoritma KNN untuk memprediksi harga emas dunia dan memperoleh kesalahan prediksi yang relatif rendah pada konfigurasi data tertentu [5]. Tinjauan literatur mengenai berbagai modifikasi

dan teknik peningkatan kinerja algoritma KNN menunjukkan bahwa performa algoritma ini sangat dipengaruhi oleh pemilihan metrik jarak, nilai K, serta strategi pembobotan tetangga [13].

Dibandingkan dengan penelitian-penelitian tersebut, studi ini berfokus pada perbandingan sistematis antara Regresi Linear dan KNN pada data harga emas harian dengan periode pengamatan yang cukup panjang dan skema pra-pemrosesan yang terstandarisasi. Penelitian ini menyoroti kondisi ketika model linier sederhana masih mampu mengungguli algoritma non-parametrik seperti KNN pada data dengan pola hubungan yang cenderung linier, serta mendokumentasikan alur eksperimen secara rinci agar dapat direplikasi dan dikembangkan oleh peneliti berikutnya.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk membandingkan kinerja dua algoritma pembelajaran mesin, yaitu Regresi Linear dan *K-Nearest Neighbors* (KNN), dalam memprediksi harga emas berdasarkan data historis. Tahapan penelitian dilakukan secara berurutan, dimulai dari studi literatur, pengumpulan dan pra-pemrosesan data, pelatihan model, evaluasi kinerja, hingga penarikan kesimpulan dan penyusunan rekomendasi. Diagram alir tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir tahapan penelitian.

2.1 Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data harga emas harian yang diperoleh dari dataset publik di platform Kaggle [14]. Dataset memuat informasi harga pembukaan, harga tertinggi, harga terendah, dan harga penutupan (*closing price*) emas internasional untuk periode Januari 2014 hingga Januari 2025. Setiap baris data merepresentasikan harga emas pada satu hari perdagangan. Variabel utama yang digunakan sebagai target (label) adalah harga penutupan emas, sedangkan variabel prediktor terdiri atas harga pembukaan, harga tertinggi, harga terendah, serta harga penutupan tertunda (*lag*) dari beberapa hari sebelumnya.

Pemilihan periode sepuluh tahun bertujuan untuk menangkap dinamika pergerakan harga emas dalam jangka menengah hingga panjang, termasuk fluktuasi yang disebabkan oleh kondisi ekonomi global, perubahan suku bunga, dan peristiwa ekonomi penting lainnya. Data yang diperoleh kemudian diolah dan dianalisis menggunakan bahasa pemrograman Python dan pustaka *machine learning* Scikit-learn [15].

2.2 Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pelatihan model. Langkah-langkah yang dilakukan meliputi:

1. Pembersihan data — menghapus baris dengan nilai hilang dan duplikasi.
2. Pembuatan fitur lag — menambahkan beberapa variabel lag harga penutupan.
3. Pembagian data latih dan data uji — 80% data sebagai latih dan 20% sebagai uji, secara berurutan berdasarkan waktu.
4. Normalisasi data — seluruh fitur numerik dinormalisasi menggunakan StandardScaler dari Scikit-learn [15] agar memiliki skala yang sebanding.

2.3 Model Regresi Linear

Regresi Linear digunakan untuk memodelkan hubungan antara variabel prediktor dengan harga penutupan emas. Secara umum, model Regresi Linear dapat dinyatakan sebagai berikut (1):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

di mana y adalah harga penutupan emas, x_1, x_2, \dots, x_p adalah variabel prediktor, $\beta_0, \beta_1, \dots, \beta_p$ adalah parameter model yang diestimasi dari data, dan ε adalah komponen galat. Model ini diimplementasikan menggunakan pustaka Scikit-learn [15] pada data latih yang telah dinormalisasi.

2.4 Model K-Nearest Neighbors (KNN)

Algoritma KNN merupakan metode non-parametrik yang melakukan prediksi berdasarkan kedekatan suatu data uji dengan sejumlah tetangga terdekatnya pada ruang fitur [13]. Pada penelitian ini, nilai k ditetapkan sebesar 5 berdasarkan beberapa percobaan awal, sehingga prediksi harga emas untuk suatu hari dihitung sebagai rata-rata harga emas dari lima data latih yang memiliki jarak terdekat.

Jarak antar titik data diukur menggunakan jarak Euclidean pada ruang fitur yang telah dinormalisasi, yang didefinisikan sebagai (2):

$$d(\mathbf{x}, \mathbf{z}) = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2 + \dots + (x_p - z_p)^2} \quad (2)$$

Proses pelatihan pada KNN bersifat implisit karena algoritma ini menyimpan seluruh data latih dan melakukan perhitungan jarak ketika prediksi dilakukan pada data uji. Implementasi algoritma KNN dalam penelitian ini juga memanfaatkan pustaka Scikit-learn [15].

Jika dibandingkan dengan penelitian yang menggunakan model *deep learning* seperti LSTM, CNN, maupun kombinasi CNN-GRU, hasil penelitian ini menunjukkan bahwa Regresi Linear masih mampu memberikan performa yang sangat baik pada data harga emas dengan pola hubungan yang cenderung linier [6], [7]. Beberapa studi tersebut melaporkan keunggulan model nonlinier ketika pola fluktuasi harga menjadi lebih kompleks atau dipengaruhi banyak variabel eksternal, sedangkan pada penelitian ini data yang digunakan relatif stabil sehingga hubungan antara fitur dan target dapat direpresentasikan secara memadai oleh model linier. Dengan demikian, pemilihan model sebaiknya mempertimbangkan karakteristik data dan tujuan analisis, bukan hanya kompleksitas algoritma yang digunakan.

2.5 Skema Pengujian dan Evaluasi

Evaluasi kinerja model dilakukan dengan menggunakan data uji yang tidak pernah digunakan selama proses pelatihan. Tiga metrik utama yang digunakan adalah koefisien determinasi (R^2), *Root Mean Squared Error*

(RMSE), dan *Mean Absolute Error* (MAE). Nilai R^2 mengukur proporsi variasi harga emas yang dapat dijelaskan oleh model, sedangkan RMSE dan MAE mengukur besarnya kesalahan prediksi dalam satuan yang sama dengan data asli. Nilai RMSE dan MAE yang lebih kecil menunjukkan kesalahan prediksi yang lebih rendah.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

dengan y_i merupakan nilai aktual, \hat{y}_i nilai prediksi, dan n jumlah data uji. Perhitungan R^2 , RMSE, dan MAE untuk masing-masing model dilakukan menggunakan fungsi evaluasi yang tersedia pada pustaka Scikit-learn.

2.6 Perangkat dan Lingkungan Eksperimen

Penelitian ini dilakukan menggunakan kombinasi perangkat lunak dan perangkat keras sebagaimana ditunjukkan pada Tabel 1. Proses pelatihan dan pengujian model dilaksanakan secara *cloud-based* pada platform Google Colab untuk memaksimalkan efisiensi komputasi serta memudahkan replikasi hasil penelitian. Implementasi model dilakukan menggunakan bahasa pemrograman Python dengan pustaka Scikit-learn untuk proses pelatihan dan evaluasi model, mengacu pada panduan implementasi *machine learning* [15].

Tabel 1. Perangkat lunak dan perangkat keras pendukung penelitian.

Hasil	Peladen	Klien
Scikit-learn	Google Colab (Cloud)	Python 3.12
NumPy, Pandas	Cloud Runtime	Windows 11
Matplotlib / Seaborn	Cloud Runtime	Linux (Ubuntu 22.04)

2.7 Desain Eksperimen dan Prosedur Replikasi

Desain eksperimen dalam penelitian ini disusun agar proses pelatihan dan evaluasi model dapat direplikasi secara konsisten. Seluruh tahapan, mulai dari pembacaan data, pra-pemrosesan, pembentukan fitur lag, pembagian data menjadi data latih dan data uji, normalisasi dengan StandardScaler, hingga pelatihan dan pengujian model, diimplementasikan dalam satu alur kerja terpadu pada lingkungan Google Colab. Parameter acak pada proses pembagian data diatur secara eksplisit sehingga peneliti lain dapat memperoleh hasil yang sebanding ketika menggunakan dataset dan konfigurasi yang sama. Transformasi normalisasi hanya dihitung berdasarkan data latih dan kemudian diterapkan ke data uji untuk menghindari kebocoran informasi (*data leakage*) yang dapat menyebabkan estimasi performa model menjadi terlalu optimistis.

Regresi Linear dilatih dengan memanfaatkan seluruh fitur yang dihasilkan dari proses pra-pemrosesan tanpa regularisasi tambahan, sehingga koefisien model dapat diinterpretasikan sebagai pengaruh relatif masing-masing fitur terhadap harga penutupan. Sementara itu, model KNN menggunakan metrik jarak Euclidean pada ruang fitur yang telah dinormalisasi. Nilai $k = 5$ dipilih berdasarkan percobaan awal yang membandingkan beberapa kandidat nilai k dengan mempertimbangkan kompromi antara akurasi, stabilitas hasil, dan kebutuhan komputasi. Pencatatan konfigurasi dan tahapan eksperimen secara eksplisit diharapkan memudahkan peneliti lain untuk melakukan penyetelan parameter lebih lanjut maupun memperluas eksperimen pada studi-studi berikutnya.

3. Hasil dan Diskusi

3.1 Statistik Deskriptif dan Korelasi Fitur

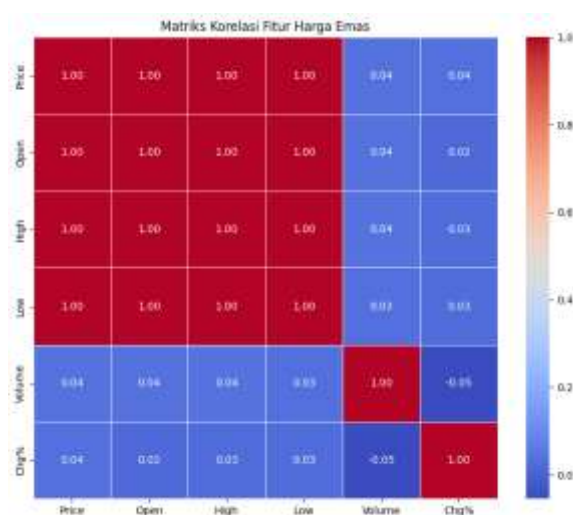
Data harga emas harian yang digunakan dalam penelitian ini merupakan subset dari dataset Kaggle “Daily Gold Price (2014-01-01 to 2025-01-06)” [14] untuk periode pengamatan yang telah dijelaskan pada Bagian 2.1. Secara umum, statistik deskriptif utama untuk variabel harga pembukaan, harga tertinggi, harga terendah, harga penutupan, volume, dan perubahan persentase harga (Chg%) menunjukkan bahwa harga emas mengalami variasi yang cukup besar selama periode pengamatan, yang mencerminkan dinamika pasar komoditas global dalam jangka

menengah hingga panjang. Rentang antara nilai minimum dan maksimum mengindikasikan adanya fluktuasi harga yang signifikan, sehingga pemodelan yang mampu menangkap pola perubahan harga menjadi penting bagi pelaku pasar. Penyebaran data yang relatif lebar juga menunjukkan perlunya pendekatan pemodelan yang mampu mengakomodasi variabilitas tersebut secara memadai. Hal ini sejalan dengan temuan beberapa penelitian sebelumnya yang melaporkan bahwa volatilitas harga emas cenderung meningkat pada periode ketidakpastian ekonomi global [2], [5].

Hubungan antar fitur ditunjukkan pada matriks korelasi yang disajikan pada Gambar 2. Terlihat bahwa variabel harga pembukaan, harga tertinggi, harga terendah, dan harga penutupan memiliki korelasi yang sangat tinggi (mendekati 1,00) satu sama lain, yang mengindikasikan adanya hubungan linier yang kuat antar variabel harga. Sebaliknya, variabel volume dan *Chg%* menunjukkan korelasi yang relatif rendah terhadap harga, sehingga kontribusinya terhadap model prediksi berbasis hubungan linier kemungkinan lebih kecil dibandingkan variabel harga utama. Informasi ini mendukung penggunaan model regresi sebagai salah satu pendekatan yang sesuai untuk memodelkan hubungan antara fitur harga dengan nilai harga penutupan emas.

Tabel 2. Hasil evaluasi model Regresi Linear dan KNN pada data uji.

Model	R ²	RMSE	MAE
Regresi Linear	0.9996	0.0021	0.0008
<i>K-Nearest Neighbors</i>	0.1107	0.0834	0.0412



Gambar 2. Matriks korelasi fitur harga emas.



Gambar 3. Perbandingan prediksi harga emas menggunakan Regresi Linear dan KNN (2021–2025).

3.2. Hasil Evaluasi Model

Kinerja model Regresi Linear dan KNN dievaluasi menggunakan metrik R^2 , RMSE, dan MAE sebagaimana telah dijelaskan pada Bagian 2.5. Ringkasan hasil evaluasi pada data uji disajikan pada Tabel 2.

Secara umum, model Regresi Linear menunjukkan performa yang jauh lebih baik dibandingkan KNN. Model Regresi Linear menghasilkan nilai R^2 sebesar **0,9996**, yang menunjukkan bahwa hampir seluruh variasi harga emas pada data uji dapat dijelaskan oleh model. Nilai RMSE dan MAE yang diperoleh masing-masing sebesar **0,0021** dan **0,0008**, sehingga tingkat kesalahan prediksi dapat dikategorikan sangat rendah.

Sebaliknya, model KNN hanya menghasilkan nilai R^2 sebesar **0,1107**, yang berarti proporsi variasi harga emas yang dapat dijelaskan oleh model sangat terbatas. Nilai RMSE dan MAE untuk KNN masing-masing sebesar **0,0834** dan **0,0412**, atau jauh lebih besar dibandingkan dengan Regresi Linear. Perbedaan ini menunjukkan bahwa, pada konfigurasi parameter yang digunakan dalam penelitian ini, KNN kurang mampu menangkap pola hubungan antara variabel prediktor dengan harga penutupan emas.

Gambar 3 memperlihatkan perbandingan antara harga emas aktual, prediksi Regresi Linear, dan prediksi K-Nearest Neighbors (KNN) pada data uji. Garis prediksi Regresi Linear tampak hampir tumpang tindih dengan garis harga aktual di sepanjang periode pengujian, yang menguatkan temuan bahwa model ini mampu merepresentasikan pola pergerakan harga emas dengan baik. Sebaliknya, prediksi KNN tampak menyimpang dan cenderung berada pada rentang yang lebih sempit, sehingga tidak mengikuti variasi harga aktual secara akurat. Pola visual ini konsisten dengan nilai R^2 , RMSE, dan MAE yang tercantum pada Tabel 2.

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa Regresi Linear memberikan performa yang sangat baik untuk kasus prediksi harga emas berbasis data historis harian, sedangkan algoritma KNN menunjukkan akurasi yang relatif rendah. Temuan ini konsisten dengan beberapa penelitian sebelumnya yang melaporkan bahwa model regresi dapat memberikan hasil yang kompetitif dalam memprediksi harga emas ketika hubungan antara variabel input dan output cenderung mendekati linier [1], [2]. Matriks korelasi pada Gambar 2 mendukung asumsi ini, karena variabel harga pembukaan, tertinggi, terendah, dan penutupan saling berkorelasi sangat tinggi.

Kinerja KNN yang kurang baik dapat disebabkan oleh beberapa faktor. Pertama, meskipun data telah dinormalisasi, pemilihan nilai k dan kombinasi fitur yang digunakan sangat memengaruhi hasil prediksi. Nilai $k = 5$ yang digunakan dalam penelitian ini mungkin belum optimal untuk struktur data yang ada. Kedua, KNN merupakan metode berbasis jarak yang sensitif terhadap distribusi dan kepadatan data pada ruang fitur [13]. Pada data dengan pola tren yang cukup jelas, seperti harga emas yang cenderung mengikuti pergerakan jangka panjang, pendekatan linier sering kali lebih stabil dan mudah beradaptasi dibandingkan metode berbasis tetangga terdekat.

Dibandingkan dengan penelitian [5] yang melaporkan bahwa KNN dapat memberikan kesalahan relatif yang rendah pada prediksi harga emas dunia, hasil penelitian ini menunjukkan bahwa performa KNN sangat bergantung pada konfigurasi fitur, pemilihan parameter, serta karakteristik dataset yang digunakan. Sementara itu, hasil yang diperoleh sejalan dengan [1], [5] yang menekankan bahwa model regresi masih relevan dan efektif untuk kasus prediksi harga emas, terutama ketika struktur hubungan antar variabel tidak terlalu kompleks.

Penelitian ini memiliki beberapa keterbatasan. Pertama, variabel prediktor yang digunakan masih terbatas pada informasi harga historis, sehingga belum mempertimbangkan faktor makroekonomi lain seperti nilai tukar, suku bunga, atau indeks saham yang pada beberapa studi terbukti berpengaruh terhadap harga emas [2], [5]. Kedua, hanya dua algoritma yang dibandingkan, yaitu Regresi Linear dan KNN. Pada penelitian selanjutnya, model lain seperti *Random Forest*, *Support Vector Regression*, atau jaringan saraf dalam dapat ditambahkan untuk memperoleh gambaran yang lebih komprehensif mengenai performa berbagai teknik *machine learning* pada prediksi harga emas.

Meskipun demikian, hasil yang diperoleh memberikan indikasi bahwa model yang relatif sederhana seperti Regresi Linear dapat menjadi pilihan yang efektif dan mudah diimplementasikan untuk prediksi harga emas berbasis data

historis, terutama pada konteks di mana interpretabilitas model dan efisiensi komputasi menjadi pertimbangan penting.

Dari sisi implementasi praktis, hasil penelitian ini menunjukkan bahwa model yang relatif sederhana seperti Regresi Linear masih sangat relevan untuk digunakan pada aplikasi pemantauan harga emas harian. Struktur model yang transparan memudahkan analisis untuk menjelaskan hasil prediksi kepada pemangku kepentingan non-teknis, karena koefisien model dapat diinterpretasikan sebagai sensitivitas harga emas terhadap perubahan fitur tertentu. Model ini juga lebih ringan secara komputasi dan lebih mudah diintegrasikan ke dalam sistem pendukung keputusan yang sudah ada. Sebaliknya, model KNN pada konfigurasi yang digunakan memerlukan sumber daya komputasi yang lebih besar pada saat prediksi dan menghasilkan pola prediksi yang kurang stabil, sehingga kurang sesuai untuk skenario yang membutuhkan respons cepat dan konsisten dalam lingkungan produksi.

4. Kesimpulan

Penelitian ini membandingkan kinerja Regresi Linear dan *K-Nearest Neighbors* (KNN) dalam memprediksi harga emas harian berdasarkan data historis yang diperoleh dari dataset Kaggle. Berdasarkan hasil analisis dan evaluasi yang telah dilakukan, Berdasarkan hasil analisis dan evaluasi, Regresi Linear memberikan performa prediksi yang sangat baik dengan nilai R^2 sebesar 0,9996 dan nilai RMSE serta MAE yang sangat rendah, sedangkan K-Nearest Neighbors (KNN) pada konfigurasi yang digunakan menunjukkan akurasi yang relatif rendah dengan nilai R^2 hanya 0,1107. Hasil ini menunjukkan bahwa KNN kurang mampu menangkap pola pergerakan harga emas pada dataset yang digunakan, sementara struktur data yang ditunjukkan oleh korelasi tinggi antar variabel harga mendukung penggunaan model linier seperti Regresi Linear untuk prediksi harga emas berbasis data historis harian.

Referensi

- [1] A. Azhari and N. Umniati, "Gold Price Prediction using Linear Regression and Polynomial Regression Method," vol. 15, no. 01, pp. 56–65, 2025.
- [2] R. Kong, "Machine Learning Models for Gold Price Prediction: A Comparative Analysis and Evaluation," *Highlights Business, Econ. Manag.*, vol. 40, pp. 429–435, 2024, doi: 10.54097/04ppnj11.
- [3] B. Xie, "Gold Price Forecast: A Summary of the Integration of Economic Factors and Calculation Methods," no. Icdse, pp. 360–365, 2025, doi: 10.5220/0013697400004670.
- [4] N. Tripurana, B. Kar, S. Chakravarty, B. K. Paikaray, and S. Satpathy, "Gold Price Prediction Using Machine Learning Techniques," *CEUR Workshop Proc.*, vol. 3283, no. May, pp. 274–281, 2021, doi: 10.54097/gdm0kc53.
- [5] A. Nanda, M. R. N. P.; Martha, Z.; Vionanda, D.; Salma, "Prediction of World Gold Price Using k-Nearest Neighbor Method," 2024, doi: doi: 10.24036/ujsds/vol2-iss4/314.
- [6] A. B. Nagata, M. S. Hidajat, D. A. Wibowo, W. Widyatmoko, and N. B. M. Yaacob, "Predicting Gold Price Movement Using Long Short-Term Memory Model," *J. Appl. Intell. Syst.*, vol. 9, no. 1, pp. 19–28, 2025, doi: 10.62411/jais.v9i1.10305.
- [7] R. A. Dalimunthe, R. T. Adek, and C. Agusniar, "Gold Price Prediction Using Long-Short Term Memory Algorithm Based on Web Application," *Int. J. Eng. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 329–339, 2025, doi: 10.52088/ijesty.v5i1.724.
- [8] G. Aditama, N. Yudistira, and W. F. Mahmudy, "Gold Prices Prediction using Univariate Long Short Term Memory Method," vol. 10, no. 2, pp. 63–73, 2025.
- [9] D. A. Adha, A. Ramadhan, H. Maulana, P. P. H. Harahap, and E. Ismanto, "Jurnal Computer Science and Information Technology (CoSciTech) Peramalan Harga Emas Berbasis Time Series Menggunakan Arsitektur LSTM Deep Learning Gold Price Forecasting Based on Time Series Using the LSTM Deep Learning Architecture," vol. 6, no. 2, pp. 329–336, 2025.
- [10] H. M. Zangana and S. R. Obeyd, "Deep Learning-based Gold Price Prediction: A Novel Approach using Time Series Analysis," *Sistemasi*, vol. 13, no. 6, p. 2581, 2024, doi: 10.32520/stmsi.v13i6.4651.
- [11] D. A. Yanumatrajaya, I. Fikri, D. Asmarayani, and H. Soetanto, "Comparing CNN and GRU for Gold Price Prediction Using Deep Learning," vol. 14, pp. 530–537, 2025.
- [12] Y. R. M. Ferdinandus, K. Kusriani, and T. Hidayat, "Gold Price Prediction Using the ARIMA and LSTM Models," *Sinkron*, vol. 8, no. 3, pp. 1255–1264, 2023, doi: 10.33395/sinkron.v8i3.12461.
- [13] R. Sakti *et al.*, "Review of Literature on Improving the KNN Algorithm," *Trans. Mach. Learn. Artif. Intell.*, vol. 11, no. 3, pp. 63–72, 2023, doi: 10.14738/tecs.113.14768.
- [14] Nisarg Chodvadiya, "Daily Gold Price (2015-2025) Time Series." [Online]. Available: <https://www.kaggle.com/datasets/nisargchodvadiya/daily-gold-price-20152021-time-series/data>
- [15] M. N. Fahmi, "Implementasi Mechine Learning menggunakan Python Library: Scikit-Learn (Supervised dan Unsupervised Learning)," *Sains Data J. Stud. Mat. dan Teknol.*, vol. 1, no. 2, pp. 87–96, 2023, doi: 10.52620/sainsdata.v1i2.31.